Introduction to Plant Breeding

A Parochial view

Origins of crops
Scientific approaches 1850... present
Plant & animal breeding compared
Achievements & questions
Matthew 7:18-7:20 A good tree cannot bring forth evil fruit, neither can a corrupt tree bring forth good fruit. Every tree that bringeth not forth good fruit is hewn down, and cast into the fire. Wherefore by their fruits ye shall know them.
The Scientific approach to plant breeding

Two strands:

1. Mendelian:
   Incorporate information from genes into selection decisions championed by plant breeders

2. Biometric:
   Incorporate information from relatives into selection decisions championed by animal breeders

Prospects: we now have the technology to combine the two.
John Goss (1824) On Variation in the Colour of Peas, occasioned by Cross Impregnation
Horticultural Transactions (Series 1) Vol:5, p. 234-237 + 1 fig
Some milestones in Mendelian genetics & breeding

1823: Knight: Dominance, recessiveness, and segregation observed in peas

1900: Rediscovery and verification of Mendel’s principles

1903: Biffen: resistance to stripe rust of wheat is Mendelian recessive.

1908: Nilsson-Ehle: seed colour in wheat is due to 3 Mendelian factors.

1923: Sax: linkage between quantitative and qualitative traits in beans.

1956: Flor: gene for gene hypothesis for host-parasite resistance

1965-70 Borlaug: Green Revolution (India & Pakistan) based on dwarfing genes.

1983: Beckmann & Soller: RFLPs for genome wide QTL detection and breeding

2001: Meuwissen *et al*: Genomic selection proposed
Dwarfing genes reduced the weight of straw, increasing:
• Nitrogen fertiliser levels.
• Higher grain yields.

Which increased susceptibility to disease. But plants were protected by newly developed:
• Fungicide

In addition, pleiotropic effects of the dwarfing gene include more grains per ear.
Information from genes.

“Brother Mendel! We grow tired of peas!”
Cartoon by J. Chase.
<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1840-50</td>
<td>de Vilmorin: progeny test in wheat, oat, and sugar-beet breeding.</td>
</tr>
<tr>
<td>1889</td>
<td>Galton: publishes Natural Inheritance, a statistical statement of the relative influence of parents</td>
</tr>
<tr>
<td>1921</td>
<td>Wright: relationships between relatives</td>
</tr>
<tr>
<td>1936</td>
<td>Smith: selection index</td>
</tr>
<tr>
<td>1947</td>
<td>Lush: Family merit &amp; individual merit as a basis for selection</td>
</tr>
<tr>
<td>1953</td>
<td>Henderson: origins of BLUP</td>
</tr>
<tr>
<td>1971</td>
<td>Patterson &amp; Thompson REML</td>
</tr>
<tr>
<td>2001</td>
<td>Meuwissen et al: Genomic selection proposed</td>
</tr>
</tbody>
</table>
Both approaches are linked by the breeders’ equation $R = h^2S$. 
Everything in plant (and animal) breeding can be judged by its effect on "the breeders’ equation."

The breeders’ equation \( R = h^2 S. \)

standardized as:

\[ R = i \ h \ \sigma_g / \ \text{time} / \ \£ \]
### Some arbitrary dates in plants breeding methods

<table>
<thead>
<tr>
<th>Year</th>
<th>Person/Sources</th>
<th>Method/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1840-50</td>
<td>de Vilmorin</td>
<td>progeny testing</td>
</tr>
<tr>
<td>1909</td>
<td>Nilsson-Ehle</td>
<td>scientific wheat breeding: pedigree breeding, bulk breeding</td>
</tr>
<tr>
<td>1878-81</td>
<td>Beal</td>
<td>corn hybrids yield more</td>
</tr>
<tr>
<td>1909</td>
<td>Shull:</td>
<td>use of F1 hybrids between inbreds in corn breeding</td>
</tr>
<tr>
<td>1924</td>
<td>Blakeslee &amp; Belling</td>
<td>report doubled haploids</td>
</tr>
<tr>
<td>1939:</td>
<td>Golden</td>
<td>single seed descent</td>
</tr>
<tr>
<td>1936</td>
<td>?</td>
<td>haploids and polyploids</td>
</tr>
<tr>
<td>Feature</td>
<td>Details</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Replicate genotypes:</td>
<td>clones</td>
<td></td>
</tr>
<tr>
<td></td>
<td>inbred lines</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DH lines</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F1 hybrids</td>
<td></td>
</tr>
<tr>
<td>Heritabilities</td>
<td>vary through replication</td>
<td></td>
</tr>
<tr>
<td>Inbreeding is quick</td>
<td>self: S1, S2, ... Sn, doubled haploids</td>
<td></td>
</tr>
<tr>
<td>Mating systems:</td>
<td>selfing, outcrossing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gms, cms, S alleles, ...</td>
<td></td>
</tr>
<tr>
<td>Polyploids:</td>
<td>haploids, allopolyploids, autopolyploids</td>
<td></td>
</tr>
<tr>
<td>Use of ancestral species</td>
<td>eg synthetic wheat</td>
<td></td>
</tr>
<tr>
<td>GxE</td>
<td>generally larger than in animals</td>
<td></td>
</tr>
<tr>
<td>Half sibs</td>
<td>have a common female parent</td>
<td></td>
</tr>
</tbody>
</table>
Methods for selection within crosses

Pedigree breeding

Single seed descent

Doubled haploids

Bulk breeding
Pedigree method

Variety A × Variety B

Bulk plot

F_1

Space planted

F_2

Plant rows

F_3

Families of plant rows

F_4

Families of plant rows

F_5

Families of plant rows

F_6

Families of plant rows

F_7

Preliminary yield trial

F_8 to F_10

Yield trials
Single Seed Descent

Variety A × Variety B

F1

Bulk plot

F2

Single plants

F3

Single plants

F4

Single plants

F5

Single plants

F6

Plant or head rows

F7

Preliminary yield trial

F8 to F10

Yield trials
Single Seed Descent

Goulden (1939)
Knott & Kumar (1975) wheat

Pedigree breeding: inbreeding & selection concur

SSD: separate inbreeding from selection (faster)

Proposed and developed for breeding.

Use in trait mapping is more recent.
Doubled Haploids

1. Variety A × Variety B
2. Bulk plot culture anthers
3. Double chromosomes with colchicine
4. F₂ Haploids
5. F₃ Plant rows
6. F₄ Plant rows
7. F₅ Preliminary yield trial
8. F₆ to F₈ Yield trials
Doubled Haploids

“The practical importance of haploids and polyploids in plant breeding is being quickly recognised and it seems possible that their artificial production will be simply a matter of technique in the near future.” Imperial Bureau of Plant Genetics, 1936

Faster than SSD

Expensive

Low efficiency in some crops

Less recombination
Bulk Breeding

Variety A × Variety B

F₁

Bulk plot

F₂

Bulk plot

F₃

Bulk plot

F₄

Bulk plot

F₅

Space planted

F₅

Plant or head rows

F₇

Preliminary yield trial

F₈ to F₁₀

Yield trials
Bulk breeding

As slow as pedigree breeding

Encourage selection in the bulk (natural & artificial)

F2s contribute unequally to inbred lines

Long history (Allard, Harlan)

Not much used in commercial plant breeding.

Regularly rediscovered by academics. And funded!
Hybrid breeding

General combining ability

Specific combining ability

Circulant partial diallels

Heterotic groups

Reciprocal recurrent selection

More money
Cereal yields in the UK

yield (t/ha)

wheat
barley
oats
winter wheat genetic and environmental trends

yield (t/ha)

first year in trial
### Linear trends in yield (t/ha)

**1982-2007 NL/RL trials**

<table>
<thead>
<tr>
<th></th>
<th>varieties</th>
<th>years</th>
</tr>
</thead>
<tbody>
<tr>
<td>winter wheat</td>
<td>0.074</td>
<td>0.010</td>
</tr>
<tr>
<td>spring barley</td>
<td>0.060</td>
<td>-0.006</td>
</tr>
<tr>
<td>winter barley</td>
<td>0.071</td>
<td>0.010</td>
</tr>
<tr>
<td>maize</td>
<td>0.109</td>
<td>0.108</td>
</tr>
<tr>
<td>sugar beet</td>
<td>0.105</td>
<td>0.112</td>
</tr>
<tr>
<td>oilseed rape</td>
<td>0.064</td>
<td>-0.019</td>
</tr>
</tbody>
</table>
N use for tillage crops: England & Wales

[Graph showing N use for tillage crops from 1960 to 2010, with kg/ha on the y-axis and years on the x-axis. The graph indicates an overall increase in N use, with a steep rise around 1980.]
Screen for sensitivity to climatic stress?
Some challenges & questions; a personal view

Have yields stopped rising?

Should we care about GxE?

What proportion of quantitative variation has originated by mutation since domestication: should we sample wild and old germplasm for yield QTL?

Do we get enough recombination?

Why are yield and quality negatively correlated?

Are the days of breeding to exploit natural variation numbered by GM?

What is the best design of a breeding programme to exploit GS?
**Why Hunt? Why Gather?**

Join the Neolithic Revolution!

- How goes the **hunt**?
  - Not so great. How’s **gathering**?
  - So-so.

- Look! A village! I wonder what they do over there...?
- Excuse me. I couldn’t help but **overhear**. Let me tell you about living the Neolithic Way!

First off – we don’t just **look around** for our food... we actually **grow** some of it ourselves, where we live!

- Plant and animal domestication is the key. We grow **edible plants** ourselves, right out of the ground, time after time!
- **Gasp**!

**Animals, too!** We **control** their reproduction to select desirable characteristics and eliminate bad ones.

- **Wow**! How can we live the Neolithic way?

You can start by **joining us** in the village! Leave your troubles behind!

- Enjoy regular meals!
- **Reshape your environment**!
- **Settle down**!

### Your KEYS to a BETTER LIFE!

- **Harness Plant Power!**
  - Learn how the seeds you crop can become next fall’s crop!
  - Use seed selection to make future plants more productive and easier to harvest!
  - Preserve and store surpluses for hard times!
  - Invent new ways of preparing and cooking plant foods!

- **Put Animals To Work For You!**
  - Learn which species are slow and submissive!
  - Use food and fences to keep them around!
  - Influence their choice of mates!
  - Breed the best and eat the rest!

**Disclaimer:** Plant and animal domestication can lead to overpopulation, deforestation, erosion, flooding, desertification, materialism, diminished nutrition, cavities, and television. Caution advised. Your results may vary.
Monday pm

• Population genetics and linkage disequilibrium
Population Genetics

Books

Felsenstein

http://evolution.genetics.washington.edu/

Weir Genetic Data Analysis 2nd ed.

http://statgen.ncsu.edu/powermarker/
GH Hardy 1877-1947

“There is no permanent place in the world for ugly mathematics.”

“I am reluctant to intrude in a discussion concerning matters of which I have no expert knowledge, and I should have expected the very simple point which I wish to make to have been familiar to biologists.”
A sufficient condition for no evolution to occur within a Mendelian population is that mutation, selection, and chance effects are all absent and that mating is at random.

The hereditary mechanism, of itself, does not change allele frequencies. The constancy of genotype frequencies then follows from the presence of random mating.
Population Genetics

The Hardy-Weinberg Law

Nothing changes except for:

- mutation
- selection
- sampling variation (drift)
- migration
- non-random mating
Population Genetics

The Hardy-Weinberg Law

<table>
<thead>
<tr>
<th>genotype</th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>X</td>
<td>2Y</td>
<td>Z</td>
</tr>
<tr>
<td>alleles</td>
<td>all A</td>
<td>½ A, ½ a</td>
<td>all a</td>
</tr>
</tbody>
</table>

Frequency of A gamete: \( X + \frac{1}{2} 2Y = p \) say
Frequency of a gamete: \( Y + \frac{1}{2} 2Y = 1-p = q \) say

with \( p + q = 1 \)

female gamete (freq)

A (p) \hspace{1cm} a (q)

male gamete (freq)

A (p) \hspace{1cm} AA (p^2) \hspace{1cm} Aa (pq)
a (q) \hspace{1cm} Aa (pq) \hspace{1cm} aa (q^2)

\( \rightarrow \)

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p^2</td>
<td>2pq</td>
<td>q^2</td>
</tr>
</tbody>
</table>

Frequency A: \( p^2 + \frac{1}{2} 2pq = p(p+q) = p \)
Polyploids

\[(p_1A_1 + p_2A_2 + p_3A_3 \ldots p_nA_n)^p\]

Eg Bufo pseudoraddei baturae
Population Genetics

Non-random mating.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p^2 + pqf$</td>
<td>$2pq(1-f)$</td>
<td>$q^2 + pqf$</td>
</tr>
</tbody>
</table>

Selfing series

<table>
<thead>
<tr>
<th>generation</th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$p^2$</td>
<td>$2pq$</td>
<td>$q^2$</td>
</tr>
<tr>
<td>1</td>
<td>$p^2 + pq/2$</td>
<td>$pq$</td>
<td>$q^2 + pq/2$</td>
</tr>
<tr>
<td>2</td>
<td>$p^2 + pq3/4pq/2$</td>
<td>$q^2 + pq3/4$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$p^2 + pq5/8pq/4$</td>
<td>$q^2 + pq5/8$</td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td>$p^2 + pq$  = $p$</td>
<td>$0$</td>
<td>$q^2 + pq$  = $q$</td>
</tr>
</tbody>
</table>
### Mixed selfing and random mating

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>$p^2 + pqf$</td>
<td>$2pq(1-f)$</td>
<td>$q^2 + pqf$</td>
</tr>
</tbody>
</table>

Just as before, but

$$f = \frac{s}{(2-s)}$$

where $s$ is the proportion of seed set by selfing or

$$f = \frac{(1-t)}{(1+t)}$$

where $t$ is the proportion of seed set by random mating
Population Genetics

Wahlund effect

Subdivided populations have reduced heterozygosity:

Frequency in population 1  \[ = \ p_1 = p+x \]
Frequency in population 2  \[ = \ p_2 = p-x \]
Average heterozygosity  \[ = \ \frac{(2p_1 q_1 + 2p_2 q_2)}{2} \]
\[ = (p+x)(1-p-x) + (p-x)(1-p+x) \]
\[ = 2pq - 2x^2 \]

Cross pops– observe excess of hets:

\[ (p+x)(1-[p-x]) + (1-p-x)(p-x) \]
\[ = 2pq + 2x^2 \]

Explanation for heterotic pools and composite varieties
Linkage Disequilibrium

Random mating between individuals generates equilibrium genotype frequencies at a single locus.
(Hardy-Weinberg equilibrium)

Random assortment of chromosomes in meiosis generates equilibrium frequencies between loci.
(Linkage equilibrium)
At equilibrium:

<table>
<thead>
<tr>
<th>loc B</th>
<th>r (B)</th>
<th>s (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loc A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p (A)</td>
<td>pr AB</td>
<td>ps Ab</td>
</tr>
<tr>
<td>q (a)</td>
<td>qr aB</td>
<td>qs ab</td>
</tr>
</tbody>
</table>

Rearranging:

<table>
<thead>
<tr>
<th>AB</th>
<th>Ab</th>
<th>aB</th>
<th>ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>pr</td>
<td>ps</td>
<td>qr</td>
<td>qs</td>
</tr>
</tbody>
</table>

Same in the next generation
With arbitrary frequencies

<table>
<thead>
<tr>
<th>Loc A</th>
<th>loc B</th>
<th>B</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>w</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>y</td>
<td>z</td>
<td></td>
</tr>
</tbody>
</table>

Compare observed and expected with $\chi^2$

<table>
<thead>
<tr>
<th>AB</th>
<th>Ab</th>
<th>aB</th>
<th>ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>w</td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>Expected</td>
<td>pr</td>
<td>ps</td>
<td>qr</td>
</tr>
</tbody>
</table>

$O - E$ | +D | -D | -D | +D |
D = observed frequency minus expected frequency

\[ D = p(AB) - p(A).p(B) \]

or

\[ -D = p(aB) - p(a).p(B) \]

etc.
Some properties of the $D$

Max value is 0.25, when $p(A)=p(B)=0.5$

At other allele freqs. max. value can be small eg

$p(A)=p(B)=0.9 \quad D_{\text{max}} = 0.09$

To make interpretation easier, define:

$D' = \frac{D}{D_{\text{max}}}$ \quad \text{range} \ 0-1$

or

$\Delta = \sqrt{p(A)p(a)p(B)p(b)} \quad \text{range} \ 0-1$
Comparison of LD measures

$\Delta \rightarrow 1$: allele freqs match, two haplotypes

$D' \rightarrow 1$: allele freqs don’t matter, three haplotypes
LD measures for multiple alleles

Calculate $D'$ or $r^2$ for each pair of alleles in turn.

Take the average, weighted by the expected frequency ($p_1p_2$)

Estimates tend to be biased upwards in small samples. The bias can be quite large.

Correct by permutation testing.
The decay of Linkage Disequilibrium

\[ D_1 = (1 - \theta) D_0 \]

\[ D_t = (1 - \theta)^t D_0 \]

<table>
<thead>
<tr>
<th># gens</th>
<th>unlinked</th>
<th>5cM</th>
<th>0.5cM</th>
<th>50k</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.50</td>
<td>0.95</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.60</td>
<td>0.95</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.01</td>
<td>0.61</td>
<td>0.95</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.61</td>
</tr>
<tr>
<td>10000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Proof

To decay, LD needs recombination. Recombination need double heterozygotes

\[
\text{AB/ab occur at a frequency } \frac{2(pr + D)(qs + D)}{2} \text{ ditto } \frac{2(ps-D)(qr-D)}{2}
\]

Arbitrarily select gamete type AB to follow over 1 generation:

\[
P_{(AB)} = \frac{2(pr + D)(qs + D)(1-\theta)}{2} + \frac{2(ps-D)(qr-D) \theta }{2}
\]

(non recs from AB/ab) (recs from Ab/aB)

Ignore terms not involving \( \theta \) to get change in \( P_{(AB)} \)

\[
= \quad [ - (pr + D)(qs +D) + (ps-D)(qr-D) ]\theta = \quad - \theta D
\]

New value of D is therefore \( D - \theta D = D(1 - \theta) \)

Over t generations: \( D_t = D_0(1 - \theta)^t \)
LD decays with time and recombination fraction
Decline in LD with genetic distance

Decline of between marker association over genetic distance. UK wheat all genomes.

LD decay human chromosome 22.

Dawson et al. 2002.
LD in Barley varieties

Chromosome 2, Barley, AGUEB SNP data
The Causes of Linkage Disequilibrium

Mutation
Sampling          drift, founder effect
Migration
Selection
Although mutation generates LD, this is not very interesting. It is the fate following mutation which is important.
Drift

\[ \mathcal{E}(\Delta^2) = \frac{1}{1 + 4N_e \theta} \]

On average, as population size and recombination increase, LD falls
Distribution of LD in founder population size 10
Migration

Pop 1 (no LD)  Pop 2 (no LD)

\[ p_1 r_1 \text{ (AB)} \quad \quad \quad \quad p_2 r_2 \text{ (AB)} \]

1:1 mix

What is the freq. of AB

Observe \[ \frac{1}{2} (p_1 r_1 + p_2 r_2) \]

Expect \[ \frac{1}{4} (p_1 + p_2)(r_1 + r_2) \]

\[ D = \frac{1}{4} (p_1 - p_2)(r_1 - r_2) \]

Zero if \[ p_1 = p_2 \] or \[ r_1 = r_2 \]
Migration – population admixture

\[ \text{av } p(A) = 0.5 \]

\[ D' \]

allele freq difference

[Graph showing the relationship between allele frequency difference and \( D' \) with a curve that increases as allele frequency difference increases.]
Hitch-hiking

Allele frequencies change at a locus as a result of selection.

As a result, closely linked polymorphisms change in frequency too.

Hitch-hiking generates LD over the whole linked region.

Is important in regions of low recombination.

These are the gene-rich regions – more opportunities for selection.
Hitch-hiking: evidence from Drosophila
Rate of recombination
An example of hitch-hiking in man.

The Morpheus gene family – function unknown – found in a class of segmental duplications.

20x normal rate of amino acid substitution.

Non synonymous substitution rate > synonymous.
Sequence alignment of two human copies of morpheus gene family.

![Graph showing nucleotide identity over bp of aligned sequence for 16 K bases.](image)
So what?

Deleterious SNPs at a high frequency are likely to be of interest.

One way they may rise in frequency is through hitch-hiking.

Therefore – look for footprints of hitch-hiking:

- High LD / low recombination / gene rich regions
- Lower heterozygosity and freq. of neutral SNPs
- Higher heterozygosity and freq. of nsSNPs
Plotting and Modelling LD

\[ E(\Delta^2) = \frac{1}{1+4N_e \theta} \]

\[ E(D') = L + (H-L)(1-\theta)^t \]
Haplotypes

Methods of determining phase:

is AaBB:

\[ AB, \ ab \]

or

\[ Ab, \ aB \]

Pedigree CEPH families

Sequencing short range

Clarke Algorithm easy to understand

EM much software - snphap

Evolutionary methods Phase